HP Proteins on Generalised Lattices

T. Vogel, M. Bachmann, and W. Janke

Institut für Theoretische Physik, Universität Leipzig, Germany

Abstract

We present investigations of HP proteins [1] in three dimensions on simple cubic and fcc lattices as well as on the triangular lattice in two dimensions.
Employing the nPERM algorithm [2] to selected sequences, we searched for ground states on these three lattices and compared the density of states and the heat capacity of sequences which are designed on the simple cubic lattice.

The HP Model

HP proteins are the simplest models of amino acid chains which imply only two types of amino acids:

- hydrophobic (H),
- polar (P).

The chain of monomers is modelled by a self-avoiding walk. The model

- is restricted to regular, e.g. simple cubic (sc) or face centered cubic (fcc), lattices,
- contains an attractive interaction between two H proteins which are neighbours on the lattice, but not covalently bound.

The total energy of a HP protein is then the sum of all H–H contacs on the lattice:

$$E = - \sum_{\binom{\text{nn-pairs } \langle i,j \rangle}{\text{not connected}}} s_i s_j \,, \qquad (1)$$

with $s_i = 1$ for hydrophobic (H) proteins and $s_i = 0$ for polar (P) ones. Figure 1 shows a HP protein with 12 amino acids (monomers) on the simple cubic lattice.

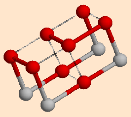


Figure 1: A HP protein consisting of 12 monomers on the simple cubic lattice. The amino acid sequence in terms of H (*red*) and P (*gray*) is HHHPHPHHPHPH. The shown conformation has the energy $E = -7$. H–H contacts are indicated by *dotted lines*.

Lattices

The most common lattice types for studies of lattice proteins are the **square** and the **simple cubic** (sc) **lattice**. There are, however, several problems with these lattice types:

- The $(n+2)$th monomer can never interact with the nth one, because it never becomes a nearest neighbour of the nth monomer with only two steps on the lattice.
- The nth monomer never interacts with the $(n+2i)$th monomer. For example, there is no conformation with an energy $E \neq 0$ for the sequence $(\mathrm{HP})_n$. These facts are unnatural lattice artefacts.
- The coordination number k, i.e., the number of nearest neighbours is rather low ($k = 4$ in two and $k = 6$ in three dimensions).

To avoid these problems, we study HP proteins on the **triangular lattice** in two dimensions ($k = 6$) and the **fcc lattice** in three dimensions ($k = 12$), respectively.

Protein Folding – Why?

The final goal:

- understanding the **folding process** and the natural, **folded conformations** of proteins. Proteins are not 'born' in their native conformation, which determines the biological function.

What we are interested in:

- **ground states** (the conformations with the lowest energy, which are assumed to be stable within normal temperature fluctuations) of some HP proteins,
- **thermodynamical properties**, e.g., the heat capacity of short, exactly treatable HP proteins.

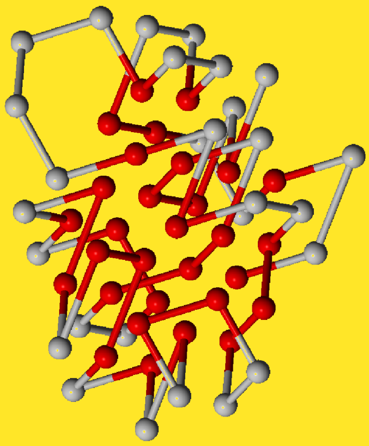


Figure 2: Lowest-energy conformation of a HP protein with 58 monomers on the fcc lattice.

Methods

1. Exact Enumeration

For short chains, we simply count exactly all possible conformations.

2. Pruned-Enriched Rosenbluth Method (PERM)

To study monomer chains with more than 20 monomers, we need a stochastic sampling algorithm, because exact enumeration becomes computationally demanding. Because of the very difficult handling of Metropolis algorithms with updates based on move sets, we use a **chain growth algorithm**.

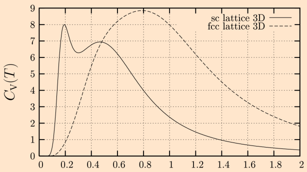
PERM [3] is such a chain growth algorithm. It combines the Rosenbluth method with the "Go with the Winners" strategy. Population control is implemented by cloning and pruning growing conformations depending on variable thresholds which are computed adaptively during the simulation.

We apply an improved PERM variant, **nPERM** [2], where the number of clones is dynamically chosen and their evolution is forced to be different.

Short HP Proteins

Here we present the result of an exact enumeration of a short HP protein with 12 monomers. Figure 1 shows the conformation with the lowest energy on the 3d sc lattice, Fig. 3 (bottom) on the 3d fcc lattice, respectively. The ground-state conformation on the 3d sc lattice is unique, i.e., the sequence is 'designing' on that lattice.

In Fig. 3 (top) the heat capacity of this protein is shown for the two lattice types. There is a clear double-peak for the 3d sc lattice. This is interpreted as a signal of a transition between the very rare low-energy states and the compact states [4].



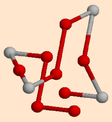


Figure 3: **Top:** The heat capacity of the mentioned 12mer on both lattices. There is a clear double peak for the sc lattice. **Bottom:** One of the numerous lowest-energy conformations of this protein on the fcc lattice

What happens with the **designingness**, when we go over to other lattice types? On the 3d fcc lattice, the measured degeneracy of the groundstate is about 20 000. There is no clearly visible double peak in the heat capacity (see Fig. 3 (top)).

Results for Longer Proteins

Finally we present the results of the ground-state search for two selected HP Proteins from [5] with 58 and 124 monomers, respectively. In Table 1 we have listed the **ground-state energies** of the two sequences found on the 2d triangular lattice, the 3d sc lattice, and the 3d fcc lattices, respectively. The energies for the 3d sc lattice are the same as in [2].

	2d triangular	3d sc	3d fcc
Sequence 58	-49	-44	-94
Sequence 124	-73	-71	-154

Table 1: Lowest energies for the selected HP proteins on different lattice types

Figure 2 shows one of the lowest-energy conformations of the sequence 58 on the fcc lattice with the energy $E = -94$.

References

[1] K.F. Lau and K.A. Dill, Macromol. **22**, 3986 (1989).

[2] H.-P. Hsu, V. Mehra, W. Nadler, and P. Grassberger, J. Chem. Phys. **118**, 444 (2002).

[3] P. Grassberger, Phys. Rev. E **56**, 3682 (1997).

[4] M. Bachmann, W. Janke, Phys. Rev. Lett. **91**, 208 105 (2003).

[5] E.E. Lattmann, K.M. Fiebig, K.A. Dill, Biochem. **33**, 6158 (1994); K.A. Dill, K.M. Fiebig, H.S. Chan, Proc. Natl. Acad. Sci. USA **90**, 1942 (1993).

This work is partially supported by the German-Israel-Foundation (GIF) under contract No. I-653-181.14/1999.

Gedruckt am Rechenzentrum der Universität Leipzig